SUPPLEMENT

A Topological View of Human CD34[+] Cell State Trajectories from Integrated Single-Cell

Output and Proteomic Data

David J.H.F. Knapp[1,2,3]§, Colin A. Hammond[1,2]§, Fangwu Wang[1,4], Nima Aghaeepour[5], Paul H. Miller[1,2], Philip A. Beer[1], Davide Pellacani[1], Michael VanInsberghe[6], Carl Hansen[6,7], Sean C. Bendall[8], Garry P. Nolan[5], Connie J. Eaves[1,2,4]*

§These authors contributed equally

1. Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, BC, Canada.
2. Department of Medicine, University of British Columbia, Vancouver, BC, Canada.
3. Weatherall Institute of Molecular Medicine, University of Oxford, UK.
4. Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada
5. Baxter Laboratory for Stem Cell Biology Department of Microbiology & Immunology, Stanford University, Palo Alto, CA, USA.
6. Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada
7. Department of Physics, University of British Columbia, Vancouver, BC, Canada.
8. Department of Pathology, Stanford University, Palo Alto, CA, USA.

Correspondence: Dr. C. J Eaves

        Terry Fox Laboratory

        BC Cancer Research Centre

        675 West 10th Avenue

        Vancouver, BC

        Canada V5Z 1L3

Phone: 604-675-8122

Fax: 604-877-0712

Email: ceaves@bccrc.ca

Running title: Analysis of differentiation trajectories

SUPPLEMENTAL METHODS

The t-SNE dimensionality reduction algorithm[36] from the R package 'Rtsne' was used to measure overall molecular similarities between analyzed events (cells). This was performed using a combination of all surface (14 parameters) and intracellular markers (26 parameters) with the perplexity parameter set to 30. All tests for modality used Hartigan's dip test in the R package 'diptest'. Population distributions were determined using 2d kernel density estimations using the function 'kde2d' in the R package 'MASS' with the default gaussian kernel and 100 bins in each t-SNE dimension with boundaries extending to 20% past the data range. Pair-wise similarities/differences between populations were determined using previously described scripts [27]. Briefly, this involved determining the construction of distance/similarity matrices based on the overlap in probability densities derived from the 2d kernel density estimates. Hierarchical clustering (the R function 'hclust') was used to provide a simplified view of such relationships given the distance matrices. Congruence among distance matrices (CADM) test were performed using the 'CADM.post' function in the R package 'ape' in order to determine whether distance matrices were consistent between conditions of interest. The CADM test was used as the null hypothesis in this case in that the matrices are dissimilar, and it can thus be used to test whether distance matrices are in fact similar.

To determine whether the inferences made from the t-SNE data were reflected in the higher dimensional data, and if other methods could have yielded an improved lineage resolution, we applied a number of different methods to a sub-sample of (up to) 100 cells from each sample/population (less where fewer analyzed cells were available). In addition to the lin-$CD34^+$ subsets, we spiked in 100 events/sample of mature lymphoid (B/T) cells with the $CD34^-$ $CD33^-CD45RA^+$ phenotype as an outgroup. On these data, pair-wise Spearman's correlation

values were generated for all cells within each population, and with all cells in each of the other populations to determine relative relationships. Similarly, we performed a hierarchical clustering (using 'hclust') between all cells based on Euclidean distances derived for the full 40 parameters. The dimensionality reduction algorithms compared included t-SNE in 2 and 3 dimensions, principle component analysis (PCA) using the R function 'princomp', Isomap in 2 and 3 dimensions using the R function 'Isomap' from the package 'RDRToolbox' with k=5, and Diffusion Map[56] with default parameters using the function 'DiffusionMap' from the R package 'destiny'[57].

In order to determine the degree of lineage information contained in different sets of protein data, we ran t-SNE on a number of subsets of protein markers using the same data subset as above. Subsets included only the TFs measured (i.e., GATA3, PAX5, PU.1, TAL1, CEBPα, and GATA1), only the surface markers measured (i.e., CD45RA, CD71, CD45, CD114, CD123, CD34, CD33, CD49f, CD10, CD135, CD38, CD90, HLADR, and CD133), all of the intracellular markers measured (i.e., pSHP2, GATA3, pCRKL, pSrc, pACC, Cyclin B1, PAX5, PU.1, pSTAT5, pAKT, pSTAT1, pSMAD2/3, pP38, pSTAT3, pMAPKAPK2, IκBα, pCREB, active β-catenin, pERK1/2, Ki67, pSykZap70, TAL1, CEBPα, pS6, GATA1, and peEF2), only the active signaling intermediates (i.e., pSHP2, pCRKL, pSrc, pACC, pSTAT5, pAKT, pSTAT1, pSMAD2/3, pP38, pSTAT3, pMAPKAPK2, IκBα, pCREB, active β-catenin, pERK1/2, pSyk/Zap70, pS6, and peEF2), or the surface markers together with the TFs. For each marker subset, density overlaps were calculated and CADM tests performed to determine whether they gave consistent relationships.

Mapping of functional progenitors to CyTOF data was performed as in[21]. Briefly, a k-nearest neighbor algorithm (the 'knnx.index' function from the R package 'FNN') was used to

determine the 10 nearest neighbors of each index-sorted progenitor within the $CD34^+$ cells from the CyTOF data using the rank-scaled (randomizing ties) overlapping markers (CD45RA, CD71, CD45, CD123, CD34, CD33, CD49f, CD10, CD135, CD38, CD90, HLA-DR, CD133). The nearest neighbors of all cells of a given type then allowed the total probability density of that cell type to be calculated. In order to assess the robustness of the obtained probability distributions we used a resampling approach. For each resampling, we took a random subset of half of the total cells with a given functional definition, and performed mapping on this subset. The degree of overlap between the sample and the overall distribution (the mapping on the total dataset) was then recorded. This was repeated 250 times.

For pseudotemporal ordering we first created gates around the highest probability area in t-SNE space for each functional progenitor type (minimum area needed to contain 15% of the total probability density) using the R function 'contourLines'. Next, for each pair along which pseudotime was to be calculated, we determined all cells in the path by creating a convex hull around the high probability cells from the two populations using the R function 'chull'. Next, a start and end location for the temporal axis were chosen based on the highest probability point for the (manually chosen) 'start' and 'end' populations. The Euclidean distance was then calculated between these two points, and from each cell to the start and end point. Cells which fell between the start and end point were retained. The Euclidean distance in t-SNE space from the start location divided by the total Euclidean distance between the start and end point was used as the pseudotemporal location. For overall summaries, pseudotime was divided into 100 bins and the median values of each marker calculated for each bin.

In order to identify more precisely how markers were changing over pseudotime, we first performed Linear, Gaussian, and Sigmoidal fits using the R function 'nlsLM' from the package

'minpack.lm'. Each fit was compared using the Bayesian information criterion (BIC) using the R

function 'BIC' and the fit with the lowest (best) score was selected. Estimations of the range of

each variable in these fits were obtained by repeating fits 1,000 times using a subset of 1% of the

total cells each time.

Supplement only references:

56. Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*. 2015;31(18):2989–2998.
57. Angerer P, Haghverdi L, Büttner M, et al. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics*. 2016;32(8):1241–1243.

SUPPLEMENTAL TABLES

Table S1. Antibodies used for index sorting of cells assayed *in vitro*

Antigen	Clone	Fluorophore
CD49f	GoH3	eFluor450
CD123	9F5	Biotin
CD45RA	HI100	eFluor605NC or Brilliant Violet 605
HLA-DR	L243	Brilliant Violet 650
CD38	HIT2	Brilliant Violet 711
CD71	OKT9	FITC
CD10	HI10a	PerCP-Cy5.5
CD135	BV10A4H2	PE
CD33	WM53	PE-CF549
CD90	5E10	PE-Cy7
CD133	AC133	APC
CD34	581	AF700
CD45	HI30	APC-eFluor780

Table S2. Antibodies used for lineage assessment of cells from the STC assays

Antigen	Clone	Fluorophore
CD45	HI30	APC-eFluor780 or Brilliant Violet 605
CD34	581	Alexa Fluor 700
CD33	WM53	PECF594
CD11b	M1/70	Brilliant Violet 711
CD14	61D3	PE-Cy7
CD15	HI98	V500
CD7	MT701	PE or FITC
CD10	HI10a	PerCP-Cy5.5
CD10	CALLA	APC
CD19	HIB19	APC-eFluor780
CD19	SJ25C1	PE
CD56	CMSSB	APC or PerCP-eFluor710
CD235a	HI264	Pacific Blue
CD1a	HI149	FITC

Table S3. Number of cells with indicated clonogenic potentials in the methylcellulose assays by input phenotype.

	BFU-E	Eos	GM-L	GM-S	Neg	Blast	GEMM
Other	5	0	7	7	16	2	9
pre-B/NK	0	0	1	5	18	0	3
LMPP	0	0	1	1	12	0	1
MLP	0	0	0	2	7	0	0
GMP	0	0	28	18	31	0	3
CMP	43	5	80	74	58	5	76
MEP	35	0	5	8	23	0	42
MPP	1	0	11	6	2	0	4
HSC	0	0	2	0	2	0	1

Table S4. Number of cells with indicated clonogenic potentials in the STC assays by input phenotype.

Clone Type	HSC	MPP	LMPP	MLP	Pre-B/NK	CMP	GMP	MEP	Other
Donor	1	1	1	1	1	1	1	1	1
Blast	0	0	0	0	0	4	0	0	0
E+NM+NK+B	0	0	0	0	0	0	0	0	0
E+NM+NK	0	0	0	0	0	2	0	0	0
E	0	0	0	0	0	9	0	6	1
NM+B+NK+T	0	1	2	0	2	7	1	0	0
NM+NK+B	0	0	0	0	0	0	2	0	0
NM+B+T	0	1	1	0	0	1	3	0	0
B+T	0	0	0	0	0	0	2	0	0
NM+B	0	0	0	0	1	1	3	0	0
B	0	0	0	0	1	1	2	0	0
NM+NK+T	0	0	0	0	0	2	1	0	0
NM+T	0	1	0	0	0	1	0	0	1
NM+NK	0	1	0	0	1	21	11	0	1
NM	0	1	0	0	3	10	11	0	1
Negative	3	7	1	7	16	55	19	9	2
Total	3	12	4	7	24	114	55	15	6

Clone Type	HSC	MPP	LMPP	MLP	Pre-B/NK	CMP	GMP	MEP	Other
Donor	2	2	2	2	2	2	2	2	2
Blast	0	0	0	0	0	2	0	0	0
E+NM+NK+B	0	0	0	0	0	1	0	0	0
E+NM+NK	0	2	0	0	0	8	0	0	0
E	0	0	0	0	0	8	1	2	0
NM+B+NK+T	0	4	1	1	1	15	1	0	1
NM+NK+B	0	0	1	0	0	1	5	0	0
NM+B+T	0	1	1	0	0	12	1	0	1
B+T	0	0	0	0	0	2	0	0	0
NM+B	0	1	0	0	0	0	2	0	0
B	0	0	0	0	2	2	1	0	0
NM+NK+T	0	0	0	0	0	1	0	0	0
NM+T	0	0	0	0	0	0	0	1	0
NM+NK	0	3	0	0	0	20	7	1	0
NM	0	0	1	0	0	12	8	0	0
Negative	1	5	2	3	4	61	22	6	1
Total	1	16	6	4	7	145	48	10	3

Table S5. Number of cells with indicated clonogenic potentials in the STC assays for each donor for new phenotypes.

Phenotype	Donor	Blast	NM+Lym+Ery	NM+Lym	NM+Ery	Ery	Lym	NM	Negative	Total
P-E	1	0	0	0	0	6	0	1	13	20
P-E	2	0	0	1	0	5	0	0	8	14
P-E	3	0	1	0	0	13	0	0	16	30
P-E	4	0	0	0	2	10	0	1	17	30
P-NM	1	0	0	5	0	0	0	5	13	23
P-NM	2	0	0	6	0	1	0	6	16	29
P-NM	3	0	0	3	1	0	1	16	9	30
P-NM	4	0	0	0	0	0	0	14	16	30
P-L	1	0	0	1	0	0	0	1	4	6
P-L	2	0	0	0	0	0	1	0	3	4
P-L	3	0	0	4	0	0	24	5	27	60
P-L	4	0	0	0	0	0	6	1	15	22
ML	1	2	1	32	0	0	3	11	21	70
ML	2	0	3	39	0	2	3	5	19	71
ML	3	5	1	24	0	0	12	10	8	60
ML	4	2	1	13	1	0	7	9	7	40
38-10+	1	0	0	0	0	0	0	0	6	6
38-10+	2	0	0	0	0	0	0	0	3	3
38-10+	3	0	0	3	0	0	7	1	49	60
38-10+	4	0	0	1	0	0	6	1	18	26
38-10-	1	0	0	9	0	0	0	1	11	21
38-10-	2	0	2	14	0	0	0	1	8	25
38-10-	3	7	2	11	1	0	9	7	23	60
38-10-	4	1	0	11	2	0	6	1	9	30

Table S6. Comparison of canonical and new phenotypes for lineage-restricted progenitors.

Phenotype	Total	Restricted Only	Other	Negative	Fisher's Exact Test p-value
P-E	94	34	6	54	0.8155706
MEP	25	8	2	15	
P-NM	112	41	17	54	0.003726553
GMP	103	19	43	41	
P-L	92	31	12	49	0.01021764
Pre-B/NK	31	3	8	20	

Table S7. Cell numbers per phenotypic population associated in each sample

	CB 1	CB 2	CB 3	CB 4	CB 5	CB 6	CB 7	CB 8
Total CD34+	4004	1266	7460	7134	4582	47947	41338	166567
pre-B/NK	43	10	182	187	84	1250	914	1110
GMP	426	80	766	988	509	6210	5756	12018
MEP	262	54	188	283	223	3957	4477	11580
CMP	435	92	1227	927	530	12356	10419	23535
MLP	230	104	454	561	321	1409	1552	5197
MPP	436	182	139	332	333	1327	1172	13797
HSC	38	12	308	134	67	884	366	1518

Samples in blue are cryopreserved, those in green were isolated fresh.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Gating hierarchies for historically defined phenotypes within the CD34$^+$ CB compartment. (**A**) Representative FACS profiles of cells isolated in the index sorting experiments. Cells pooled from 3 methylcellulose culture experiments are shown. Channels were mean-scaled per experiment to allow pooling. (**B, C**) Gating hierarchy for mass cytometry data shown on a random sampling of 100,000 cell events. Axes represent the asinh (marker intensity/5) value for the indicated channel. Canonical phenotypes are shown in (**B**) and new phenotypes in (**C**)

Figure S2. Gating hierarchies for assessment of lineage outputs in the STC assay.
Representative FACS profiles of clonal analyses that combined contain examples of each investigated mature cell population. Specific classification of each clone is listed in the top-right corner of each set of plots. Axes represent the asinh (marker intensity/5) value for the indicated channel.

Figure S3. Comparison between measurements of protein and transcript levels across all phenotypes. Each point indicates a pair-wise correlation between the ranked intensity value for a CyTOF-determined protein measurement, and the ranked transcript level in the same phenotype obtained from GSE42414 for 7 of the phenotypic subsets of CD34$^+$ CB cells analyzed in both studies. The overall mean of all pair-wise comparisons per probe/antibody pair are shown as blue lines. Probe/antibody pairs where median pair-wise correlations are statistically different from 0

(Holm-corrected p ≤0.05) are placed along horizontal green lines, while those that do not reach significance are placed along black lines.

Figure S4. Relative levels of intracellular proteins in different phenotypically defined subsets of CD34$^+$ CB cells. Histograms show the asinh (marker intensity/5) for each phenotypically defined population. Lines indicate median values. Lack of significance (Holm-corrected p>0.05) by Kruskal-Wallis rank sum tests between population medians across samples is indicated by an "NS" in the top right of the panel with unmarked panels being significant (Holm-corrected p ≤0.05).

Figure S5. The CMP phenotype is a mixture of cell types with distinct molecular characteristics. (**A**) Gating used to define CMP subpopulations. (**B**) Histograms indicating the asinh (marker intensity/5) of a selection of analyzed transcription factors and signaling molecules in the major CMP sub-populations identified. (**C**) Histograms indicating the asinh (marker intensity/5) of all surface features analyzed in each of the major CMP subpopulations identified. Lines indicate median values.

Figure S6. Comparison between dimensionality reduction methods. To compare different methods of dimensionality reduction and cell comparison, up to 100 cells from each phenotypically defined population were sampled from each subset. This included cells from the CD34$^-$CD33$^-$CD45RA$^+$ population (mature lymphoid cells) as an out group. Each cell type is color coded. (**A**) Pair-wise Spearman's correlation between cells within and across phenotypes given all 40 parameters. (**B**) Hierarchical clustering (using the Ward method) based on Euclidean

distance given all 40 parameters. (**C**) t-SNE dimensionality reduction in 2 and (**D**) 3 dimensions. (**E**) PCA showing the first 3 principal components. (**F**) Isomap based dimensionality reduction based on the 5 nearest neighbors to 2, and (**G**) 3 dimensions. (**H**) DiffusionMap dimensionality reduction to 3 dimensions.

Figure S7. Functional mapping to molecular data yields generally reproducible profiles. (**S**) Highest probability density interval (containing 15% of the total mapping probability) for each functional capability defined in STC assays. (**B, C**) Reproducibility analyses for mapping functionally assessed cells back to mass cytometric measurements. For each run, a random sample of half the total cells with each functional category were taken and used to generate a 2-dimensional distribution (using the same k-means mapping as was used in the full analyses). The degree of overlap between this and the reference distribution (the overall dataset) was then recorded. This was repeated a total of 250 times to generate the likely reproducibility of each distribution. (**B**) The mapping of progenitor types determined in the methylcellulose assays. (**C**) Reproducibility of progenitor types as measured in the STC assays. In both cases the overall number of cells in each category are listed beneath the category label.

Figure S8. Inferred molecular differences between functionally defined progenitor types. (**A, C**) Histograms showing the asinh (marker intensity/5) of the 5 most significantly differential surface (upper) and intracellular (lower) markers in the nearest neighbors for each (**A**) progenitor type measured in the methylcellulose assays and (**C**) progenitor type measured in the STC assays. Median values are displayed as thick lines. (**B, D**) Statistical significance for each mark between the nearest neighbors of each (**B**) progenitor type measured in methylcellulose assays or

13

(**D**) STC assays, expressed as the $-\log_{10}$ (Holm-corrected p-value) of the Kruskal-Wallis rank sum test. Significant values (Holm-corrected p-value ≤ 0.05) are shown as solid circles, and those below significance are shown as hollow circles. A solid line shows the 0.05 threshold.

Figure S9. Additional relative marker intensities of the nearest neighbors for each progenitor cell type detected in the methylcellulose assays. All surface (**A**) or intracellular (**B**) markers that were not in the top 5 are shown (significance testing and top 5 marks are shown in Figure S8). Differences that did not reach significance are marked with "NS". Lines indicate median values.

Figure S10. Additional relative marker intensities of the nearest neighbors for each progenitor cell type detected in the STC assays. All surface (**A**) or intracellular (**B**) markers that were not in the top 5 are shown (significance testing and top 5 marks are shown in Figure S8). Differences that did not reach significance are marked with "NS". Lines indicate median values.

Figure S11. Additional marker changes over pseudotime for the transition from STC-initiating cells that made only CD34$^+$ cells to STC-E. Pseudotime, cell color, and fits are as is indicated in Figure 6. Markers are ordered by the amount of change over pseudotime (residuals of the median bins).
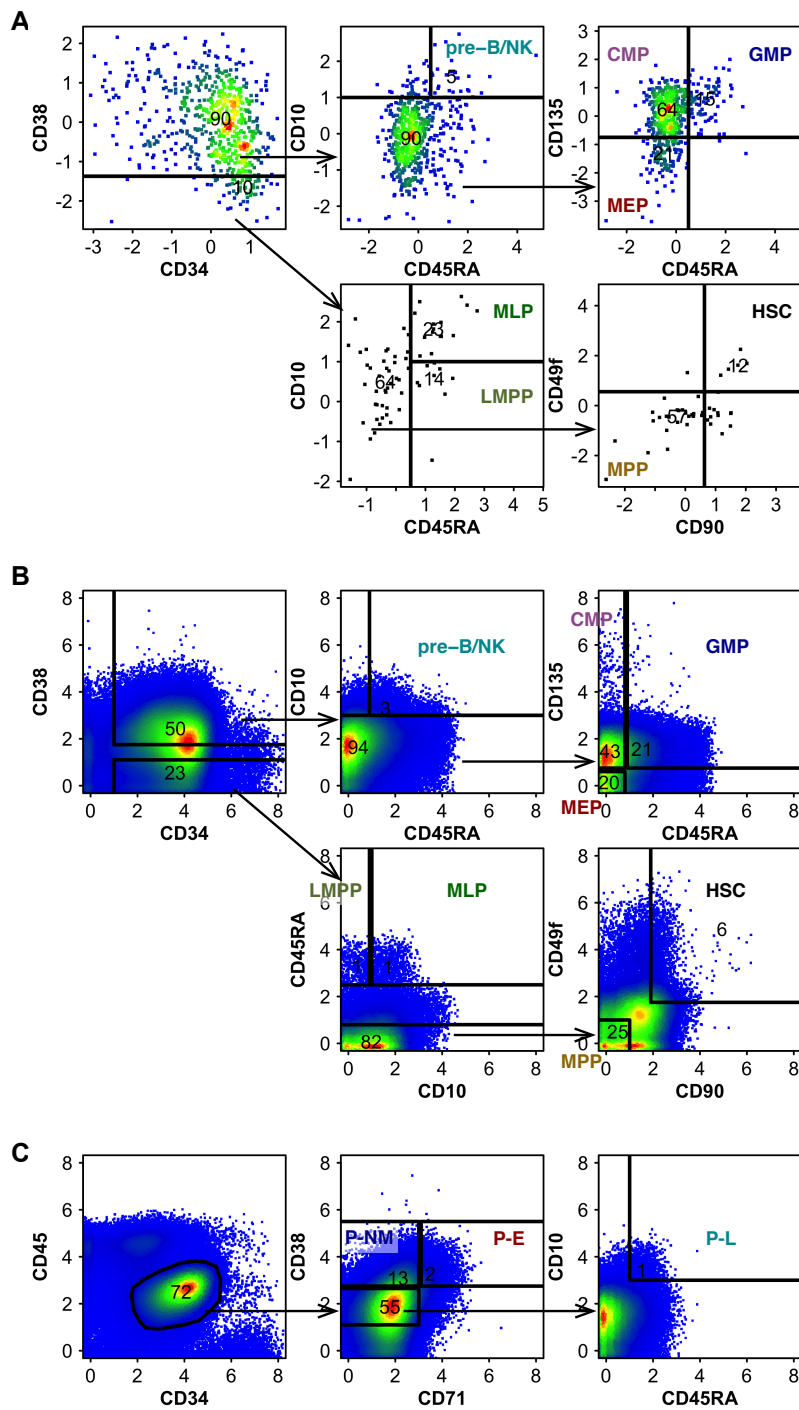
Figure S12. Additional marker changes over pseudotime for the transition from STC-initiating cells that made only CD34$^+$ cells to STC-NM. Pseudotime, cell color, and fits are as

is indicated in Figure 6. Markers are ordered by the amount of change over pseudotime (residuals of the median bins).

Figure S13. Additional marker changes over pseudotime for the transition from STC-initiating cells that made only CD34+ cells to STC-B. Pseudotime, cell color, and fits are as is indicated in Figure 6. Markers are ordered by the amount of change over pseudotime (residuals of the median bins).

Figure S1

Figure S1. Gating hierarchies for historically defined phenotypes within the CD34+ CB compartment. (**A**) Representative FACS profiles of cells isolated in the index sorting experiments. Cells pooled from 3 methylcellulose culture experiments are shown. Channels were mean-scaled per experiment to allow pooling. (**B**, **C**) Gating hierarchy for mass cytometry data shown on a random sampling of 100,000 cell events. Axes represent the asinh (marker intensity/5) value for the indicated channel. Canonical phenotypes are shown in (**B**) and new phenotypes in (**C**).

Figure S2

Figure S2. Gating hierarchies for assessment of lineage outputs in the STC assay. Representative FACS profiles of clonal analyses that combined contain examples of each investigated mature cell population. Specific classification of each clone is listed in the top-right corner of each set of plots. Axes represent the asinh (marker intensity/5) value for the indicated channel.

Figure S3

Figure S3. Comparison between measurements of protein and transcript levels across all phenotypes. Each point indicates a pair-wise correlation between the ranked intensity value for a CyTOF-determined protein measurement, and the ranked transcript level in the same phenotype obtained from GSE42414 for 7 of the phenotypic subsets of CD34+ CB cells analyzed in both studies. The overall mean of all pair-wise comparisons per probe/antibody pair are shown as blue lines. Probe/antibody pairs where median pair-wise correlations are statistically different from 0 (Holm-corrected p ≤0.05) are placed along horizontal green lines, while those that do not reach significance are placed along black lines.

Figure S4

Figure S4. Relative levels of intracellular proteins in different phenotypically defined subsets of CD34+ CB cells. Histograms show the asinh (marker intensity/5) for each phenotypically defined population. Lines indicate median values. Lack of significance (Holm-corrected p>0.05) by Kruskal-Wallis rank sum tests between population medians across samples is indicated by an "NS" in the top right of the panel with unmarked panels being significant (Holm-corrected p ≤0.05).
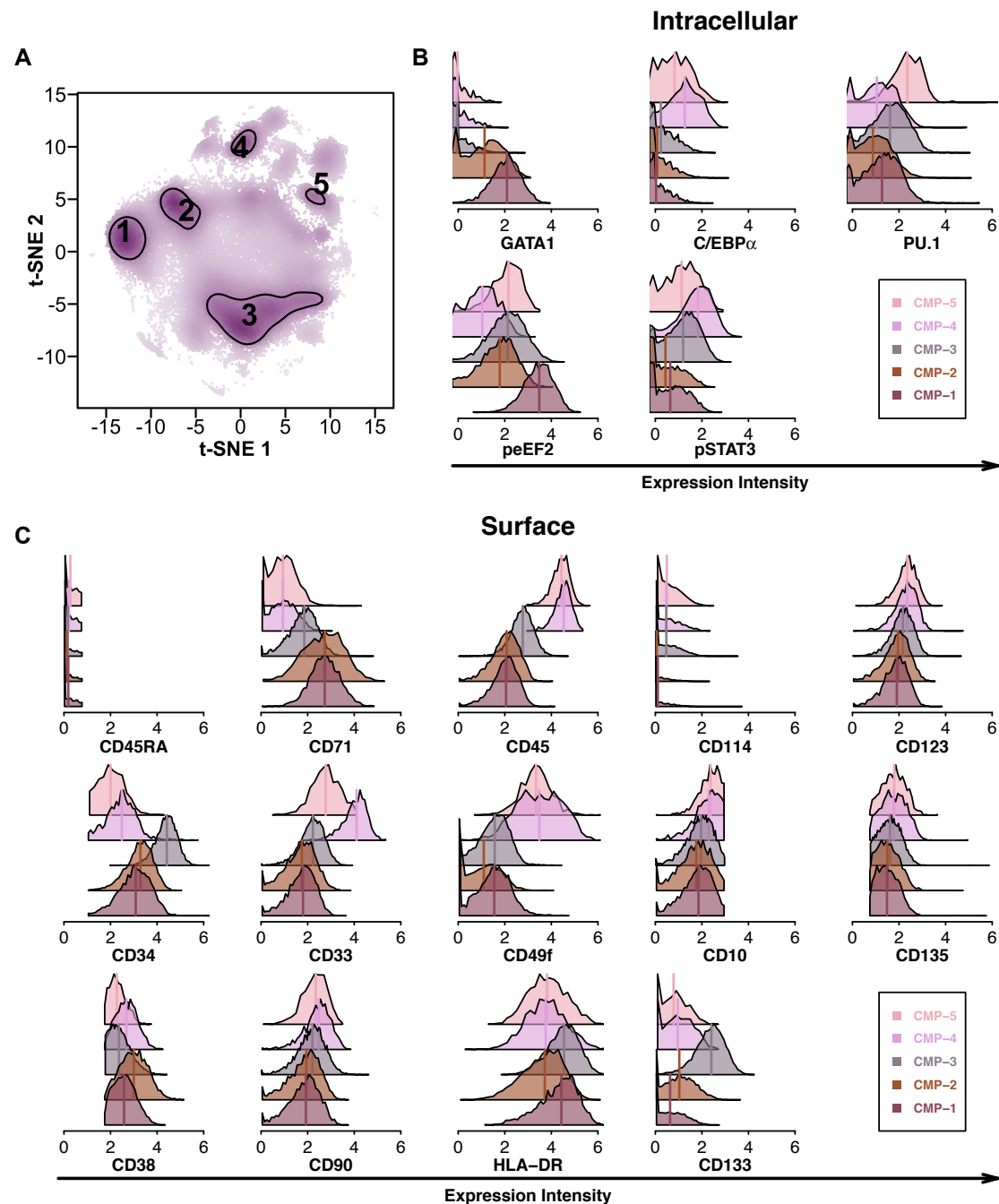
Figure S5. The CMP phenotype is a mixture of cell types with distinct molecular characteristics. (**A**) Gating used to define CMP subpopulations. (**B**) Histograms indicating the asinh (marker intensity/5) of a selection of analyzed transcription factors and signaling molecules in the major CMP sub-populations identified. (**C**) Histograms indicating the asinh (marker intensity/5) of all surface features analyzed in each of the major CMP subpopulations identified. Lines indicate median values.

Figure S6

Figure S6. Comparison between dimensionality reduction methods. To compare different methods of dimensionality reduction and cell comparison, up to 100 cells from each phenotypically defined population were sampled from each subset. This included cells from the CD34⁻CD33⁻CD45RA⁺ population (mature lymphoid cells) as an out group. Each cell type is color coded. (**A**) Pair-wise Spearman's correlation between cells within and across phenotypes given all 40 parameters. (**B**) Hierarchical clustering (using the Ward method) based on Euclidean distance given all 40 parameters. (**C**) t-SNE dimensionality reduction in 2 and (**D**) 3 dimensions. (**E**) PCA showing the first 3 principal components. (**F**) Isomap based dimensionality reduction based on the 5 nearest neighbors to 2, and (**G**) 3 dimensions. (**H**) DiffusionMap dimensionality reduction to 3 dimensions.

Figure S7

Figure S7. Functional mapping to molecular data yields generally reproducible profiles. (**A**) Highest probability density interval (containing 15% of the total mapping probability) for each functional capability defined in STC assays. (**B, C**) Reproducibility analyses for mapping functionally assessed cells back to mass cytometric measurements. For each run, a random sample of half the total cells with each functional category were taken and used to generate a 2-dimensional distribution (using the same k-means mapping as was used in the full analyses). The degree of overlap between this and the reference distribution (the overall dataset) was then recorded. This was repeated a total of 250 times to generate the likely reproducibility of each distribution. (**B**) The mapping of progenitor types determined in the methylcellulose assays. (**C**) Reproducibility of progenitor types as measured in the STC assays. In both cases the overall number of cells in each category are listed beneath the category label.

Figure S8

Figure S8. Inferred molecular differences between functionally defined progenitor types. (**A, C**) Histograms showing the asinh (marker intensity/5) of the 5 most significantly differential surface (upper) and intracellular (lower) markers in the nearest neighbors for each (**A**) progenitor type measured in the methylcellulose assays and (**C**) progenitor type measured in the STC assays. Median values are displayed as thick lines. (**B, D**) Statistical significance for each mark between the nearest neighbors of each (**B**) progenitor type measured in methylcellulose assays or (**D**) STC assays, expressed as the -$\log_{10}$ (Holm-corrected p-value) of the Kruskal-Wallis rank sum test. Significant values (Holm-corrected p-value ≤ 0.05) are shown as solid circles, and those below significance are shown as hollow circles. A solid line shows the 0.05 threshold.

Figure S9. Additional relative marker intensities of the nearest neighbors for each progenitor cell type detected in the methylcellulose assays. All surface (**A**) or intracellular (**B**) markers that were not in the top 5 are shown (significance testing and top 5 marks are shown in Figure S8). Differences that did not reach significance are marked with "NS". Lines indicate median values.
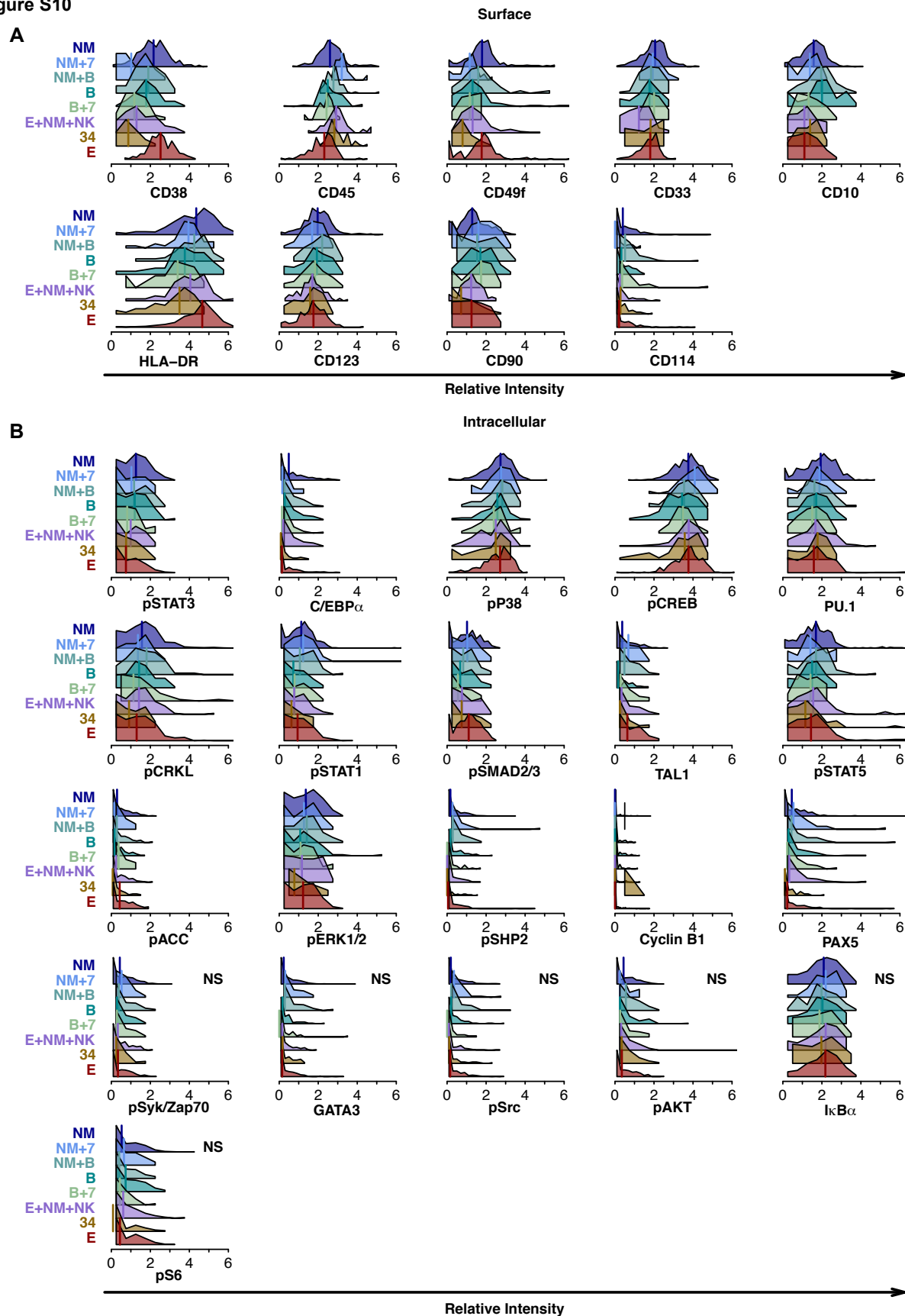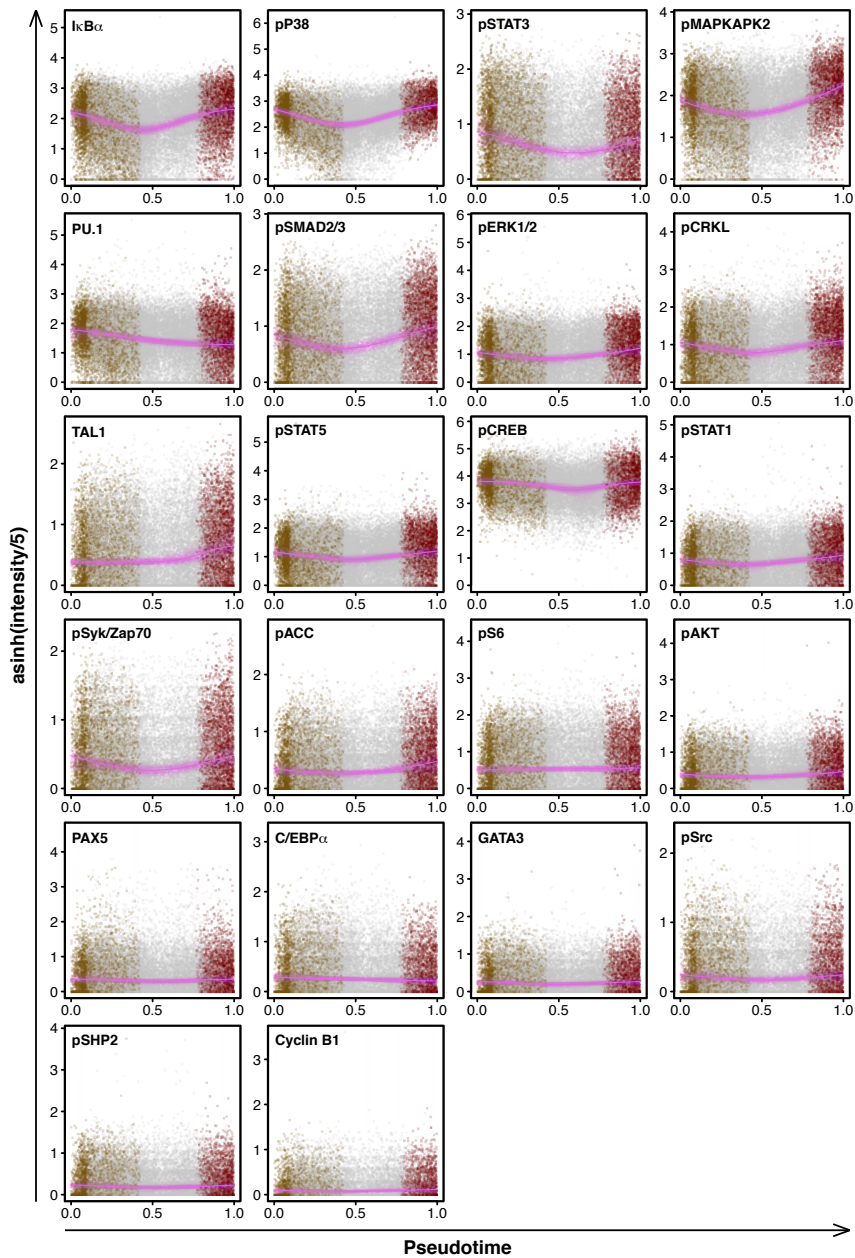
Figure S10. Additional relative marker intensities of the nearest neighbors for each progenitor cell type detected in the STC assays.
All surface (**A**) or intracellular (**B**) markers that were not in the top 5 are shown (significance testing and top 5 marks are shown in Figure S8). Differences that did not reach significance are marked with "NS". Lines indicate median values.

Figure 11

Figure S11. Additional marker changes over pseudotime for the transition from STC-initiating cells that made only CD34⁺ cells to STC-E. Pseudotime, cell color, and fits are as is indicated in Figure 6. Markers are ordered by the amount of change over pseudotime (residuals of the median bins).
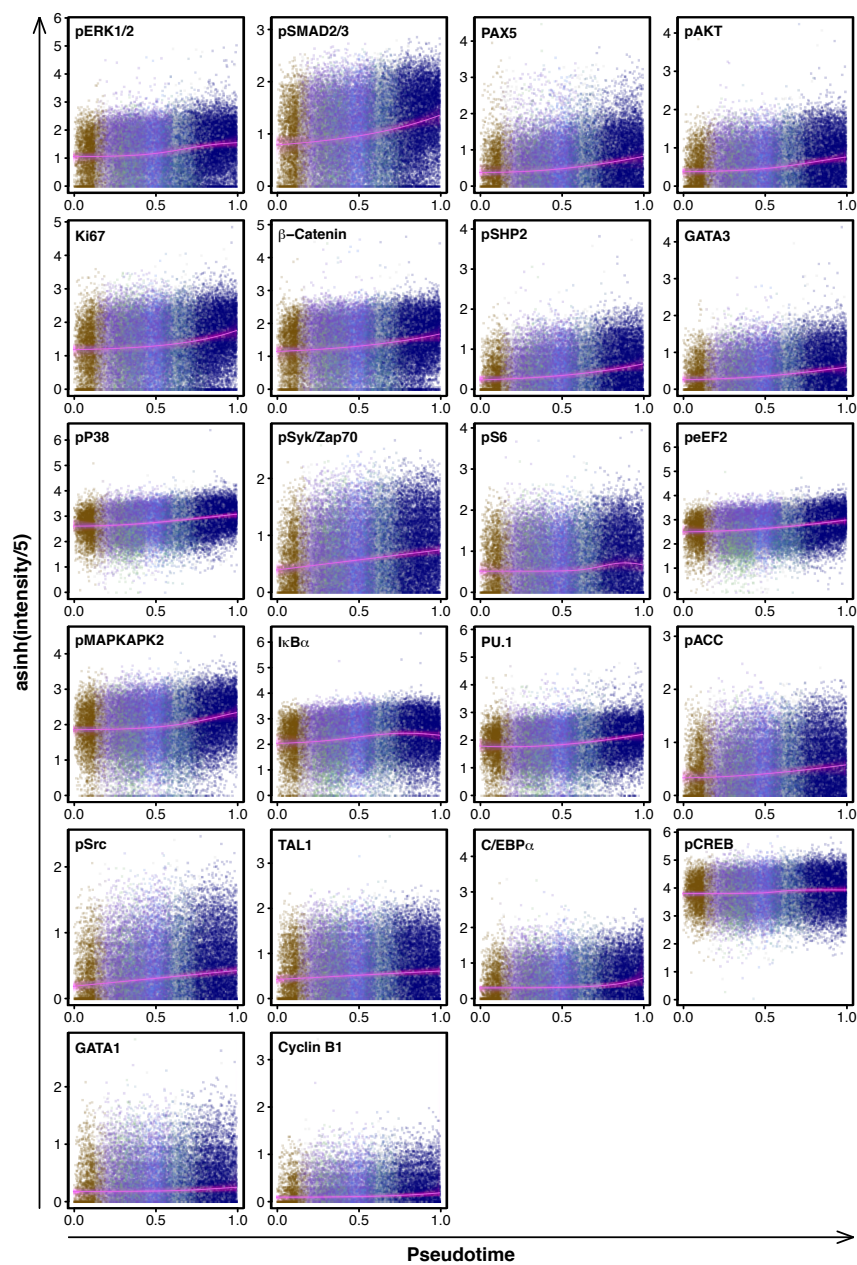
Figure S12

Figure S12. Additional marker changes over pseudotime for the transition from STC-initiating cells that made only CD34⁺ cells to STC-NM. Pseudotime, cell color, and fits are as is indicated in Figure 6. Markers are ordered by the amount of change over pseudotime (residuals of the median bins).
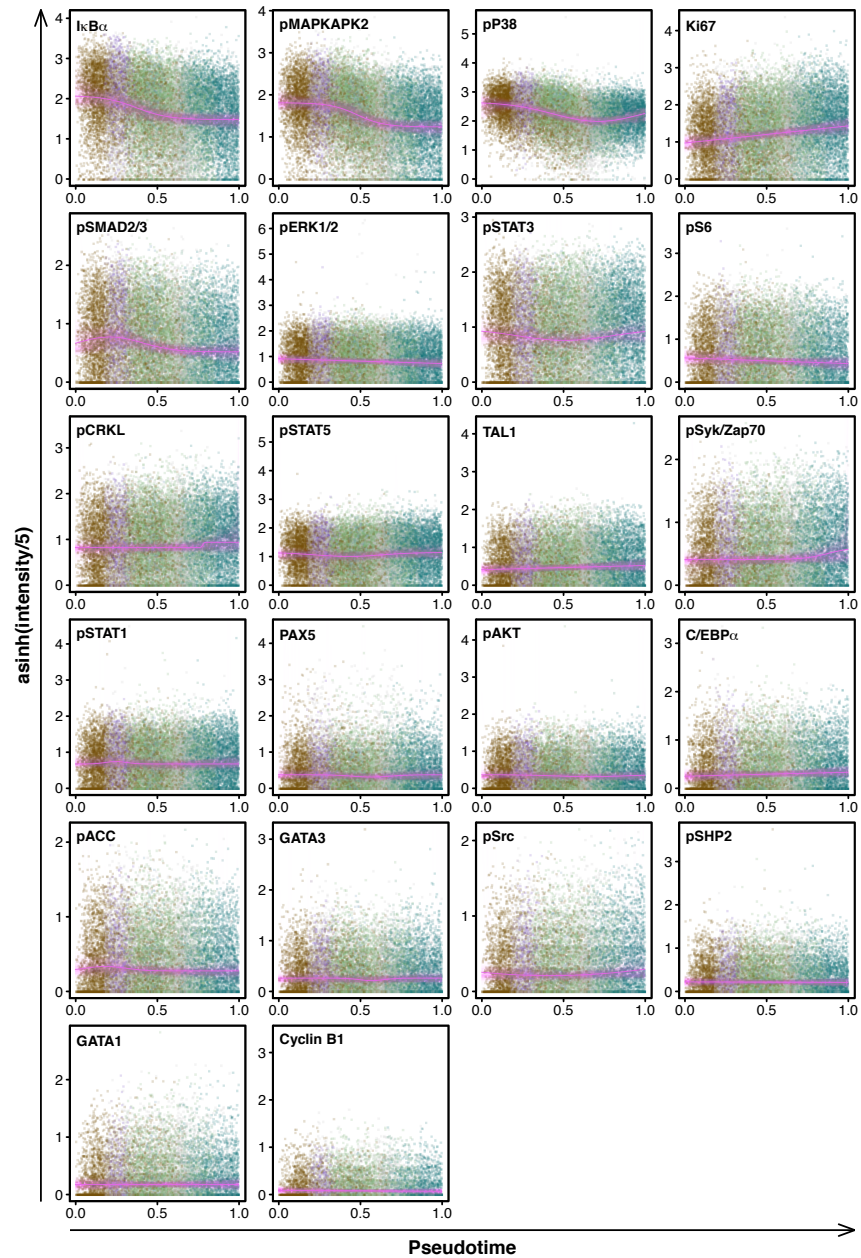
Figure S13

Figure S13. Additional marker changes over pseudotime for the transition from STC-initiating cells that made only CD34+ cells to STC-B. Pseudotime, cell color, and fits are as is indicated in Figure 6. Markers are ordered by the amount of change over pseudotime (residuals of the median bins).